


RESEARCH ARTICLE

Dynamic effects of prognostic factors and individual survival prediction for amyotrophic lateral sclerosis disease

Baoyi Huang, Xiang Geng, Zhiyin Yu, Chengfeng Zhang & Zheng Chen 

Department of Biostatistics, School of Public Health (Guangdong Provincial Key Laboratory of Tropical Disease Research), Southern Medical University, Guangzhou, China

Correspondence

Zheng Chen, Department of Biostatistics, School of Public Health, Southern Medical University, Tonghe, Guangzhou 510515, China. Tel: +862061649462; Fax: +862061648319; E-mail: zheng-chen@hotmail.com

Funding Information

This work was supported by the National Natural Science Foundation of China (Grant numbers [82173622], [81673268], and [82073675]), and the Guangdong Basic and Applied Basic Research Foundation (Grant numbers [2022A1515011525]).

Received: 27 January 2023; Revised: 22 March 2023; Accepted: 23 March 2023

Annals of Clinical and Translational Neurology 2023; 10(6): 892–903

doi: 10.1002/acn3.51771

Abstract

Objective: Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease affecting motor neurons, with broad heterogeneity in disease progression and survival in different patients. Therefore, an accurate prediction model will be crucial to implement timely interventions and prolong patient survival time. **Methods:** A total of 1260 ALS patients from the PRO-ACT database were included in the analysis. Their demographics, clinical variables, and death reports were included. We constructed an ALS dynamic Cox model through the landmarking approach. The predictive performance of the model at different landmark time points was evaluated by calculating the area under the curve (AUC) and Brier score. **Results:** Three baseline covariates and seven time-dependent covariates were selected to construct the ALS dynamic Cox model. For better prognostic analysis, this model identified dynamic effects of treatment, albumin, creatinine, calcium, hematocrit, and hemoglobin. Its prediction performance (at all landmark time points, $AUC \geq 0.70$ and Brier score ≤ 0.12) was better than that of the traditional Cox model, and it predicted the dynamic 6-month survival probability according to the longitudinal information of individual patients. **Interpretation:** We developed an ALS dynamic Cox model with ALS longitudinal clinical trial datasets as the inputs. This model can not only capture the dynamic prognostic effect of both baseline and longitudinal covariates but also make individual survival predictions in real time, which are valuable for improving the prognosis of ALS patients and providing a reference for clinicians to make clinical decisions.

Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease associated with the loss of brain and spinal cord motor neurons, causing paralysis of voluntary muscles, which may lead to problems in action, speech, and breathing.¹ Paralysis of muscles, respiratory failure, and loss of vital functions are common causes of death. Once ALS symptoms onset, the patient's quality of life will be greatly reduced, and along with the increasing mortality risk, this will also bring a tremendous burden to the social health care system.²

Precision medicine³ is committed to improving all aspects of population health care. One aspect of precision medicine is to build a predictive model with good

performance that can provide accurate survival prediction and effective prognostic analysis. ALS has a widely heterogeneous clinical presentation, disease progression, and ultimately survival. The typical time from symptom onset to death is 2–3 years,⁴ but 5%–20% of patients may progress very slowly and even survive for decades.⁵ Consequently, both the prediction of survival at the individual level and the identification of biomarkers for prognostic analysis are challenging.⁶

Although existing medicines are approved for ALS treatment,^{7,8} there remains a significant unmet need to prevent muscle function decline and prolong patient survival.⁹ After the diagnosis of ALS, patients may experience symptom attacks at any time; thus, long-term follow-up is needed to monitor their condition. At each follow-up

visit, the prediction of the patient's future survival probability (or mortality risk) and the dynamic prognostic changes in biomarkers can provide some reference for clinicians to make clinical decisions, such as controlling the medication dosages. Clinicians also need tools to predict the timing of interventions. Accurate predictive models will be critical in improving the efficiency of therapeutics and prolonging patient survival.

During the long-term follow-up of ALS patients, demographic information is collected, as are various laboratory indicators, vital signs, and respiratory capacity at each follow-up. Therefore, most ALS datasets include both baseline (static) and time-dependent covariates.¹⁰ For example, the patient's age at diagnosis, sex, and treatment are recorded upon entry into the study and are among the baseline covariates. At the same time, albumin, creatinine and calcium in the blood are measured at multiple follow-ups. The values of these covariates may change with the patient's condition, making them longitudinal time-dependent covariates.

Most studies on ALS have focused on the disease progression and survival of patients,^{11,12} and most have used traditional Cox proportional hazards (PH) models to evaluate specific prognostic indicators.^{12–16} These models are usually constructed based on only the baseline covariates and use a constant hazard ratio (HR) value (at baseline) to assess the differences between groups. As shown in Fig. 1A, these models often make predictions at baseline ($s_0 = 0$), which are referred to as the static (baseline) prediction model. However, during the actual follow-up process, especially long-term follow-up, the effects of both time-dependent covariates and baseline covariates may change over time, which are called dynamic prognostic effects, so the HR is a value that may not stay constant. Existing static prediction models may have limited ability to deal with these covariates. On the one hand, the application of the Cox model must satisfy the PH assumption; on the other hand, the static prediction model fails to fully utilize longitudinal information to predict disease progression and survival in real time.

Therefore, we wanted to introduce a dynamic prediction model for patients with ALS. As shown in Fig. 1B, the landmarking approach is a common dynamic prediction method^{17–19} that considers patients still at risk at different landmark time points s_l and the longitudinal information collected up to the prediction time (both baseline and later time points). The Cox model combined with the landmarking approach¹⁸ can capture the dynamic prognostic effect of both baseline covariates and time-dependent covariates but also make full use of the longitudinal information of ALS patients to dynamically predict their survival probability.

Here, we aimed to construct an ALS dynamic survival prediction model by inputting clinical trial data from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database into a dynamic Cox regression with a landmarking approach. The prognostic analysis of this model can track the dynamically changing effects of different covariates in clinical follow-up; this model can make predictions of the future survival probability of ALS patients at the time point of interest. Thus, this prediction model can provide a reference for clinicians to make clinical decisions.

Methods

Data source

Data used in this study were obtained from the PRO-ACT repository.^{1,20} The PRO-ACT dataset includes more than 10,000 clinical patient records involved in 23 phase II/III clinical trials. The latest version of the PRO-ACT database (1st August 2022) was used in this work.

The data cleaning process was performed as described in Appendix S1 and Fig. S2.

Statistical methods

Landmarking approach

The basic idea of the landmarking approach is to choose a series of meaningful time points in advance and evaluate the status of each patient at each time. Using the landmarking approach to construct a dynamic prediction model, we should first determine the maximum prediction time of interest s_{\max} , which is usually based on the maximum follow-up time in the data. As mentioned in the Introduction, the follow-up time of ALS patients from onset to death is generally 2–3 years.⁴ The maximum survival time in our dataset was 37.7 months (more than 3 years) after data cleaning. Therefore, we defined s_{\max} as 18 months, which was the intermediate time point of the maximum follow-up interval, and the expected prediction interval was $[0, 18]$. In addition, 18 months is the longest recorded stable period for ALS patients.²¹ Thus, the follow-up information of all the patients during the previous 18 months could be included in the analysis, with fewer than 20% of the patients having their follow-up information after 18 months ignored. We selected a prediction window w of 6 months to predict patient survival in the next half-year. During the period from baseline ($s_0 = 0$) to 18 months ($s_L = s_{\max} = 18$), we selected 19 landmark time points (one per month) for equally spaced follow-up times. At each landmark time point s_l ($l = 0, 1, \dots, 18$), we selected the patients still at risk (alive and

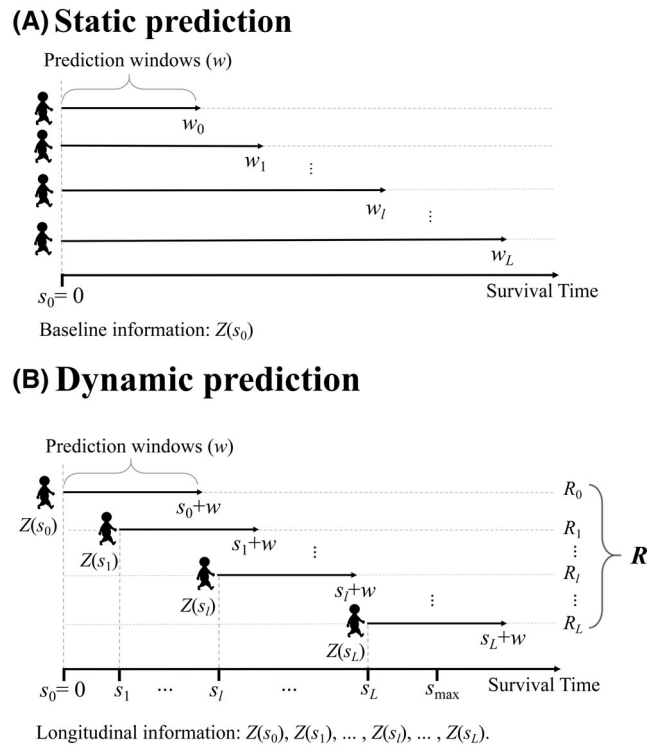


Figure 1. Static prediction and dynamic prediction.

being followed up) to constitute the corresponding landmark dataset R_l ($l=0, 1, \dots, 18$), as shown in Fig. 1B. Ignoring the events that occurred after $s_l + w$ months, which meant those during each period $[s_l, s_l + w]$, the survival status was recorded as censored if the patient was still alive.

Dynamic Cox prediction model

Through the above landmarking approach, a “super prediction dataset” R was formed by stacking all 19 landmark datasets R_l together. The dynamic Cox model was constructed in R by adding the normalization of landmark time points $t_{LM} = s_l / (s_L - s_0)$ as the function $\theta(s_l) = \theta_1 t_{LM} + \theta_2 t_{LM}^2$. We included $\beta_k(s_l) = \beta_{k,0} + \beta_{k,1} t_{LM} + \beta_{k,2} t_{LM}^2$ as a time-varying effect function for the k th covariate Z_k at prediction time s_l , where $k \in \{1, 2, \dots, K\}$, K being the total number of covariates; and $\beta_{k,0}, \beta_{k,1}, \beta_{k,2}$ represent the parameters of the constant, linear and quadratic terms, respectively. The following model estimates the dynamic risk rate at any time point t from s_l to $s_l + w$:

$$h(t, Z|s_l) = h_0(t) \times \exp[\theta(s_l)] \times \exp\left[\sum_{k=1}^K Z_k^T(s_l) \cdot \beta_k(s_l)\right],$$

$$s_l \leq t \leq s_l + w$$

Model evaluation

We also constructed a static Cox model based on the baseline data, which only made predictions at baseline, while the dynamic Cox model could predict at each landmark time point s_l . To show the difference in predictive ability between the model that only used baseline data and the dynamic prediction model that used both baseline and longitudinal data, we constructed a conditional Cox model²² based on baseline data at each landmark time point, whose predictive performance we compared with that of the dynamic Cox model.

We performed a Monte Carlo cross-validation simulation in the model evaluation. The data were randomly divided into a training set (70% of the sample) and a test set (30% of the sample) for cross-validation to avoid overfitting. The area under the curve (AUC) is typically used for discrimination.²³ If the AUC is higher, the consistency is better, which means that the predicted survival probability is more consistent with the patient’s actual survival time. The Brier score is a measure of calibration, and it equals the squared difference between the probability predicted by a model and the observed result.²⁴ The smaller the Brier score is, the higher the accuracy of the model prediction. The average AUC and Brier score values were calculated at each landmark time point.

Results

Description of predictors

In the end, 1260 patients were included in the analysis, with a censoring rate of 29.76%. Their survival time ranged from 0.13 to 37.73 months. As shown in Table 1, we performed a data description of 37 covariates at baseline, which included 4 baseline covariates and 33 time-dependent covariates. For example, the mean value of the continuous variable age was 57.84 years, with a

standard deviation of 10.24, and the binary variable treatment had 875 patients in the active group and the rest in the placebo group. To avoid the small effect of each additional unit value of continuous variables, we preprocessed some covariates, such as dividing the age variable by 10, so the one-unit effect could be explained by the risk increase in patients every 10 years. We used the baseline data to construct a univariate Cox model and found that 17 covariates, such as age, treatment and albumin, all had a significant effect on the survival outcome ($p < 0.05$). The PH assumption

Table 1. The results of the descriptive analysis of baseline data.

Variables (unit) (reference item)	Mean \pm SD (N = 1260)	Univariate analysis		<i>p</i> value	PH test
		Coef	SE (coef)		
Age (10 year)	5.78 \pm 1.02	0.315	0.036	<0.001	0.493
Sex = Female (ref: male)	510	0.064	0.068	0.351	0.940
Race = Caucasian (ref: other)	1214	0.283	0.199	0.156	0.460
Treatment = Active (ref: placebo)	875	−0.224	0.072	0.002	<0.001
Used riluzole = Yes (ref: no)	812	0.133	0.071	0.060	0.457
Albumin (10 g/L)	4.44 \pm 0.30	−0.648	0.111	<0.001	0.600
ALT/SGPT > 40 (ref: 0–40 U/L)	365	−0.063	0.075	0.401	0.898
AST/SGOT > 40 (ref: 0–40 U/L)	200	0.034	0.092	0.715	0.104
Bicarbonate (10 mmol/L)	2.63 \pm 0.32	1.114	0.111	<0.001	0.499
Bilirubin total (μ mol/L)	10.14 \pm 4.95	0.021	0.006	<0.001	0.223
BUN (mmol/L)	5.71 \pm 1.87	0.044	0.020	0.026	0.040
Calcium (mmol/L)	2.36 \pm 0.16	−0.636	0.165	<0.001	<0.001
Chloride (mmol/L)	10.24 \pm 0.29	−0.032	0.122	0.793	0.047
Creatinine (10 μ mol/L)	6.59 \pm 1.76	−0.060	0.020	0.003	0.001
Glucose >5.5 (ref: \leq 5.5 mmol/L)	475	0.215	0.069	0.002	0.441
Hematocrit (10%)	4.34 \pm 0.38	−0.205	0.090	0.023	0.293
Hemoglobin (10 g/L)	14.39 \pm 1.30	0.002	0.027	0.935	0.565
Potassium (mmol/L)	4.18 \pm 0.33	−0.040	0.107	0.705	0.067
Sodium (10 mmol/L)	13.92 \pm 0.24	−0.297	0.146	0.041	0.137
RBC (10E9/L)	4.72 \pm 0.46	−0.232	0.076	0.002	0.349
WBC (10E9/L)	6.83 \pm 1.87	0.020	0.018	0.251	0.216
BPD (10 mmHg)	8.06 \pm 1.05	0.085	0.033	0.011	0.761
BPS (10 mmHg)	13.16 \pm 1.72	0.069	0.019	<0.001	0.049
Pulse (10 beats/min)	7.69 \pm 1.18	0.078	0.028	0.005	0.302
Resp rate (breaths/min)	17.45 \pm 3.16	0.035	0.011	0.002	0.089
Q1_Speech (score)	2.76 \pm 1.26	0.004	0.027	0.873	0.323
Q2_Salivation (score)	3.08 \pm 1.15	0.024	0.030	0.410	0.948
Q3_Swallowing (score)	3.10 \pm 1.06	0.027	0.031	0.382	0.742
Q4_Handwriting (score)	2.40 \pm 1.39	0.005	0.025	0.844	0.408
Q5_Cutting (score)	2.06 \pm 1.41	0.009	0.024	0.702	0.350
Q6_Dressing_and_Hygiene (score)	1.92 \pm 1.27	0.018	0.027	0.513	0.316
Q7_Turning_in_Bed (score)	2.41 \pm 1.28	0.002	0.027	0.933	0.132
Q8_Walking (score)	2.26 \pm 1.12	−0.012	0.031	0.698	0.288
Q9_Climbing_Stairs (score)	1.40 \pm 1.41	0.020	0.024	0.417	0.212
Q10_Respiratory (score)	3.47 \pm 0.76	0.111	0.046	0.016	0.888
ALSFERS total (10 score)	2.49 \pm 0.73	0.041	0.047	0.377	0.330
Forced vital capacity 1 (L)	2.57 \pm 1.10	0.031	0.032	0.327	0.943
Global PH test					<0.001

BPD, blood pressure diastolic; BPS, blood pressure systolic; BUN, blood urea nitrogen; PH, proportional hazards; RBC, red blood cells; Resp rate, respiratory rate; WBC, white blood cell.

test showed that six covariates did not meet the PH assumption at baseline, and the global model outcome was the same ($p < 0.05$). The time-dependent covariates were collected from repeated multiple measurements, whose values usually change over time and which may not satisfy the PH assumption at later time points.

Predictor selection

The cleaned dataset retained 37 covariates, only some of which would necessarily be clinically or statistically significant in the model in actual clinical practice. Therefore, we needed to filter the covariates to ensure the predictive performance of the final model by selecting as few but statistically significant covariates as possible.

Due to the linear relationship between ALSFRS Total and Q1–Q10, we removed ALSFRS Total first, with the remaining 36 covariates as Dataset 1. After that, we used ALSFRS Total to replace Q1–Q10, with the remaining 27 covariates becoming Dataset 2. Based on Dataset 1 and Dataset 2, we added their covariates and the interaction terms with time to construct the respective dynamic Cox models.

ALS dynamic prognostic and prediction model

We added the covariates in Dataset 1 and their interaction terms with time to the initial dynamic Cox model for stepwise backward regression (satisfying $p < 0.05$) and finally selected 19 covariates as Dataset 3. The prognostic results of the dynamic Cox models based on Dataset 3 are listed in Table S1, and the dynamic HRs of covariates are shown in Fig. S1. Based on these results, we finally selected 10 covariates with obvious time-varying effects and clinical significance as Dataset 4: treatment, albumin, calcium, creatinine, hematocrit, hemoglobin, age, sex, bicarbonate and glucose. Thus, based on Dataset 4, we constructed one static Cox model and one ALS dynamic Cox model (final model).

Dynamic prognostic analysis

The results of the static Cox model and the ALS dynamic Cox model based on Dataset 4 are presented in Table 2, and the dynamic HRs of the covariates in the final model are shown in Fig. 2.

In the static Cox model, being in the active treatment group was a protective factor, and the risk was 0.774

Table 2. Static Cox model and ALS dynamic Cox model based on Dataset 4.

	Static Cox model				ALS dynamic Cox model			
Variables (unit)	Coef		SE (Coef)	<i>p</i> value	Coef		SE (Coef)	<i>p</i> value
Age (10 year)	β_1	0.309	0.039	<0.001	$\beta_{1,0}$	0.243	0.017	<0.001
Sex = Female	β_2	−0.058	0.091	0.523	$\beta_{2,0}$	−0.210	0.037	0.016
Trea = Active	β_3	−0.256	0.073	<0.001	$\beta_{3,0}$	−1.407	0.109	<0.001
					$\beta_{3,1}$	4.254	0.466	<0.001
					$\beta_{3,2}$	−3.144	0.445	<0.001
Albu (10 g/L)	β_4	−0.263	0.122	0.031	$\beta_{4,0}$	−1.272	0.127	<0.001
					$\beta_{4,1}$	0.882	0.217	0.030
Bica (10 mmol/L)	β_5	0.906	0.113	<0.001	$\beta_{5,0}$	1.014	0.040	<0.001
Calc (mmol/L)	β_6	−0.409	0.221	0.064	$\beta_{6,0}$	2.645	0.318	<0.001
					$\beta_{6,1}$	−4.291	0.561	<0.001
Crea (10 μmol/L)	β_7	−0.101	0.023	<0.001	$\beta_{7,0}$	−0.294	0.021	<0.001
					$\beta_{7,1}$	0.214	0.036	0.004
Gluc >5.5	β_8	0.137	0.069	0.048	$\beta_{8,0}$	0.235	0.029	<0.001
Hema (10%)	β_9	−1.067	0.252	<0.001	$\beta_{9,0}$	0.099	0.244	0.832
					$\beta_{9,1}$	−3.047	0.413	<0.001
Hemo (10 g/L)	β_{10}	0.388	0.078	<0.001	$\beta_{10,0}$	0.219	0.076	0.123
					$\beta_{10,1}$	0.579	0.127	0.014
t_{LM}					θ_1	6.441	1.461	0.022
$(t_{LM})^2$					θ_2	2.285	0.444	<0.001

Albu, albumin; Bica, bicarbonate; Calc, calcium; Crea, creatinine; Gluc, glucose >5.5 mmol/L (ref: ≤ 5.5 mmol/L); Hema, hematocrit; Hemo, hemoglobin; Sex = Female (ref: male); Trea, treatment = Active (ref: placebo); $t_{LM} = s_l / (s_l - s_0)$ is the normalization of landmark time points; $\theta(s_l) = \theta_1 t_{LM} + \theta_2 t_{LM}^2$ is the time function; $\beta_k(s_l) = \beta_{k,0} + \beta_{k,1} t_{LM} + \beta_{k,2} t_{LM}^2$ is the time-varying effect function for the k th covariates Z_k . The k in $\{1, 2, \dots, K\}$, K is the total number of covariates; the $\beta_{k,0}, \beta_{k,1}, \beta_{k,2}$ separately represents the parameters of the constant, linear, and quadratic terms.

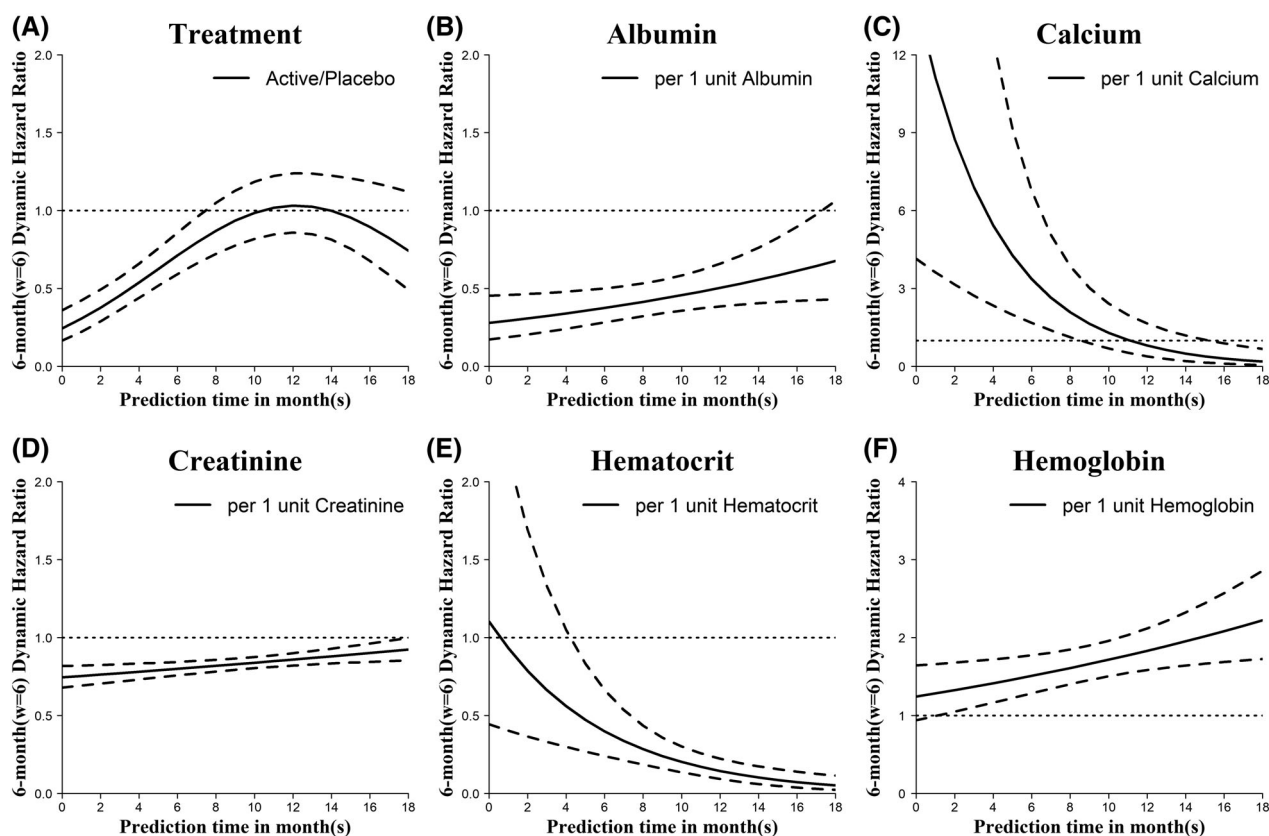


Figure 2. Time-varying effect for covariates.

times that in the placebo group ($HR = \exp(-0.256) = 0.774$). Albumin ($HR = 0.769$ per 10 g/L increase) and creatinine ($HR = 0.904$ for every 10 mmol/L increase) were also protective factors that reduced the mortality risk in disease progression. Hematocrit was a protective factor ($HR = 0.344$ per 10% increase), whereas hemoglobin was a danger factor ($HR = 1.474$ per 10 g/L increase). Although calcium was a protective factor without statistical significance in static Cox models ($p = 0.064$), it does not satisfy the PH assumption, so the results obtained by traditional Cox model analysis may be unreliable.

As shown in Fig. 2A, the HR of treatment (= active group) was 0.245 at baseline, and the effect gradually decreased with the change in follow-up time in the final model. After the seventh month of follow-up, the 95% confidence interval of the HR value included 1. In Fig. 2B–D, both albumin and creatinine were indeed protective factors at baseline, but both with time-varying effects decreased as prediction time increased, with the 95% confidence interval of dynamic HR approaching 1 at follow-up around Month 18. Neither hematocrit nor

hemoglobin was statistically significant at the beginning of follow-up (Fig. 2E,F). As the prediction time changed, hematocrit gradually became a strong protective factor ($HR = 0.052$); in the meantime, hemoglobin became a danger factor, conferring a risk of 2.221 per 10 g/L increase. The higher the calcium value at the beginning of follow-up, the higher the mortality risk (Fig. 2C), making it a danger factor. However, as the prediction time changed (follow-up progress), the dynamic HR value of calcium gradually decreased, the 95% confidence interval included 1 at 9 months of follow-up, and it became a protective factor after nearly 16 months of follow-up.

Finally, the remaining covariates age, bicarbonate, and glucose (>5.5 mmol/L) have all been confirmed as danger factors^{25–27} that shorten patient survival time. For example, bicarbonate had $HR = 2.474$ in the static Cox model, and in the final model, there was 2.757 times the risk for every 10 mmol/L bicarbonate increase. In addition, sex (=female), although not statistically significant in the static Cox model, was found to be a protective factor in the final model ($HR = 0.811$), which is also in line with previous findings.²⁸

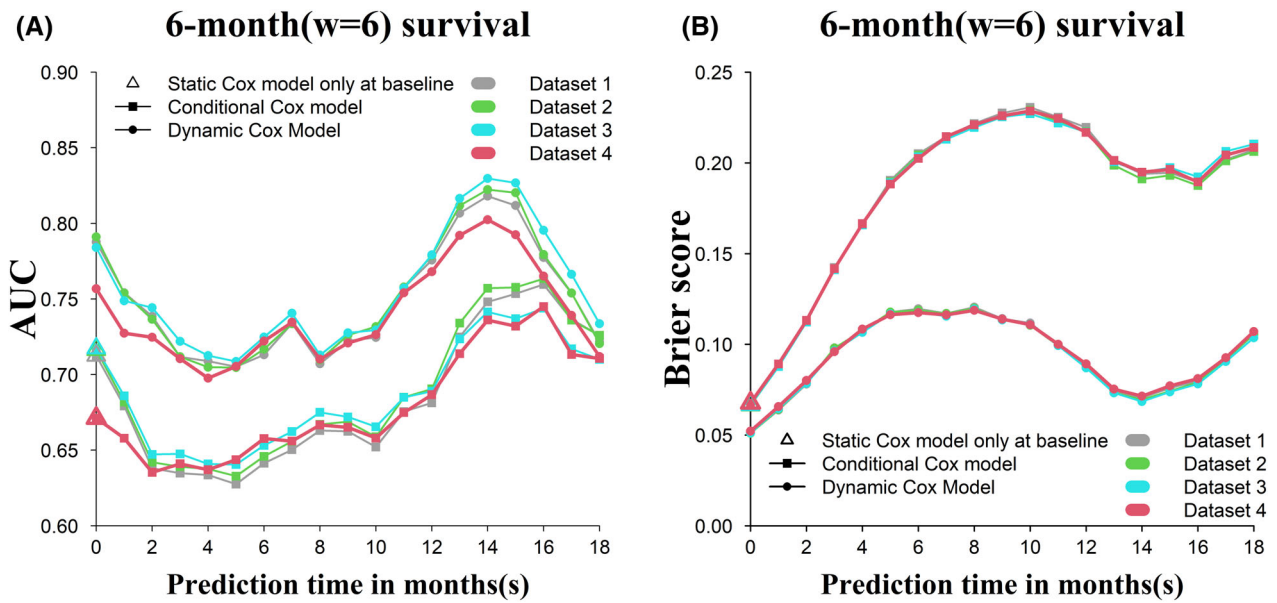


Figure 3. The results of model evaluation.

Predictive performance assessment

We performed internal validation by Monte Carlo cross-validation, separately examining the predictive performance of both the static Cox model and the dynamic Cox models built based on the four datasets. As shown in Fig. 3, the time-dependent AUC and Brier score of the dynamic Cox model outperformed those of the static Cox model and the conditional Cox model at baseline. At the subsequent landmark time points, the dynamic Cox model also outperformed the conditional Cox model. The static Cox model could only make predictions at baseline, while the dynamic Cox model could make predictions at different prediction time points; thus, the two models could not be directly compared. Figure 3 also shows that from Dataset 1 to Dataset 4, although the model construction became more streamlined, the overall time-dependent AUC was greater than 0.70, and the Brier score stayed below 0.12. The predictive performance was not reduced by the reduction in covariates, and the performance was relatively stable.

Individual survival prediction

The main purpose of constructing a dynamic prediction model was to provide dynamic and real-time survival predictions for ALS patients, which would be valuable in clinical decision-making because the outcome determines who should receive early treatment and when it should be applied. The ALS dynamic Cox model (final model) was as follows:

$$\begin{aligned} \hat{h}(t|Z, s_l) = & \hat{h}_0(t) \times \exp [6.441 \cdot t_{LM} + 2.285 \cdot t_{LM}^2] \\ & \times \exp [Age(s_l) \cdot 0.243 + Sex(s_l) \cdot (-0.210) \\ & + Bica(s_l) \cdot 1.014 + Gluc(s_l) \cdot 0.235 \\ & + Trea(s_l) \cdot (-1.407 + 4.254 \cdot t_{LM} - 3.144 \cdot t_{LM}^2) \\ & + Albu(s_l) \cdot (-1.272 + 0.882 \cdot t_{LM}) \\ & + Calc(s_l) \cdot (2.645 - 4.291 \cdot t_{LM}) \\ & + Crea(s_l) \cdot (-0.294 + 0.214 \cdot t_{LM}) \\ & + Hema(s_l) \cdot (0.099 - 3.047 \cdot t_{LM}) \\ & + Hemo(s_l) \cdot (0.219 + 0.579 \cdot t_{LM})], \\ & s_l < t < s_l + w \end{aligned}$$

where $\hat{h}_0(t)$ is the baseline hazard rate estimated by the dynamic Cox model, $t_{LM} = s_l / (s_L - s_0)$ are the normalized landmark time points; items such as $Age(s_l)$ are the values of the covariates at the landmark time point s_l ; and $\hat{h}(t|Z, s_l)$ could be calculated by inputting the covariate value of patient i at the landmark time point s_l . Through the following formula, the conditional survival probability of patient i ($i \in \{1, 2, \dots, n\}$, where n is the sample size) at landmark time point s_l can be obtained:

$$\hat{S}_i(s_l + 6|s_l, Z_i) = \exp \left[- \int_{s_l}^{s_l+6} \hat{h}(t|Z_i, s_l) dt \right]$$

We selected two patients with different degrees of disease progression who had the specific covariate information given in Table S2. As shown in Table S2 and Fig. 4A, patient 1 was a man in the placebo group with poor

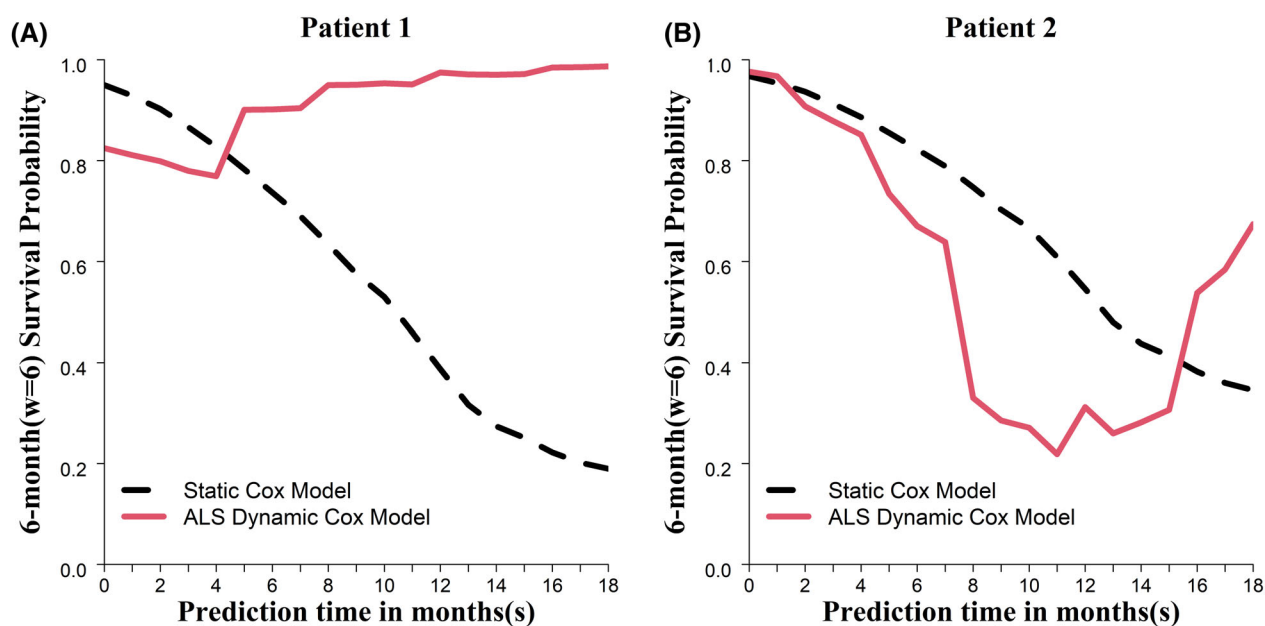


Figure 4. Dynamic individual survival predictions.

performance in the remaining covariates at baseline, and the ALS dynamic Cox model predicted that his 6-month survival probability was approximately 0.80. As the follow-up progressed, albumin, creatinine, and hematocrit were increased, bicarbonate was the danger factor, and glucose and hemoglobin were decreased. Under the combined effect of these covariates, the 6-month survival probability remained higher than 0.90. Conversely, although patient 2 (Fig. 4B) was a female in the active group and the other covariates performed well at baseline, the predicted 6-month survival probability was higher than that of patient 1 at baseline (approximately 0.98). However, during the follow-up from 4 to 8 months, the patients' condition changed; the protective factors albumin, creatinine, and hematocrit decreased; and the danger factors bicarbonate, calcium, and hemoglobin increased, so the 6-month survival probability of patient 2 decreased to 0.28. As the follow-up went on, the corresponding covariates improved after 12 months, so the 6-month survival probability increased.

At the same time, we calculated the survival probability predicted by the static Cox model. The static Cox model could only use the baseline covariates to predict the survival probability at baseline, which was continuously decreasing and could not reflect the real condition changes of patients in real time.

Discussion

Dynamic prediction is a method that uses updated follow-up information to accurately predict the survival/

risk of patients with a certain disease.²⁹ With the development of precision medicine, the diagnosis and treatment of ALS diseases need to enter the era of individualized medicine,^{2,6} and clinicians also need tools to predict the timing of treatment interventions, so prediction models will be crucial to improve the effectiveness of treatment.³⁰ The purpose of this paper was to construct an ALS dynamic Cox prognosis and prediction model based on longitudinal ALS clinical trial data from the PRO-ACT database.

The ALS dynamic Cox prognostic model accounts for the time-varying effects of the longitudinal time-dependent covariates. First, we found that the effect of treatment gradually decreased with follow-up time, and the dynamic HR was not statistically significant after 7 months of follow-up. Just as the patients' long-term drug administration process may lead to the emergence of drug resistance, the efficacy of the treatment decreases with time.³¹ Second, ALS patients generally have lower levels of creatinine than healthy people³²; 5% albumin for plasmapheresis has an acceptable safety profile in ALS patients.³³ Therefore, elevated creatinine and albumin are protective factors that reduce the risk of death and improve the survival probability as the disease progresses.^{1,34} Our model found that creatinine and albumin were indeed protective factors at baseline. It was also found that the effects of creatinine and albumin gradually decreased as the follow-up progressed.

In addition, the results of dynamic prognostic analysis revealed some new findings. Calcium dysregulation in the blood plays a central role in the pathophysiology of

ALS.³⁵ Our final model found that calcium was a danger factor at baseline, but in the late follow-up (16 months later), calcium gradually became a protective factor. Both hemoglobin and hematocrit were found to be danger factors at baseline, consistent with earlier findings,^{27,36} but without statistical significance in our model. As the prediction time increased, hemoglobin became a strong danger factor, and hematocrit transformed into a protective factor. In the existing clinical studies, the conclusions about the impact of calcium, hematocrit, and hemoglobin on the long-term survival of ALS patients are not comprehensive, so the roles of these three variables in the pathophysiology of ALS deserve further research. We hope that the results of dynamic prognostic analysis will provide a reference for clinical researchers when studying the long-term survival of ALS patients.

The ALS dynamic Cox prediction model we constructed can estimate the patients' conditional survival probability at different prediction time points, which can provide some reference for clinicians to make clinical decisions. Taking patient 1 as an example, at $s_l = 6$, the predicted 6-month survival probability is over 0.90, which indicates that the patient can live for at least 6 months. The patient's survival probability stayed high during the middle and late follow-up, also indicating that the plan of therapy made by the clinicians was appropriate and prolonged the patient's survival time. Clinicians can decide to stay on the current course or even reduce the dosage of a drug. In contrast, for patient 2, the survival probability declined to approximately 0.28 during the 4- to 8-month follow-up period. During this 4-month span, when a decreasing survival probability was predicted, clinicians needed to identify the cause of the patient's condition changes in a timely manner and make decisions to intervene. Therefore, the survival probability could be used as an alternative indicator to predict changes in patients' conditions in advance. Within the continued follow-up of patient 2, the predicted 6-month survival probability increased after 12 months. At this time, it is possible that the plan of therapy was modified by the clinicians, perhaps to a more suitable plan for the patient, thereby improving the patient's condition.

Many studies based on ALS datasets also take ALS disease progression (functional impairment) as the outcome variable of interest, and most of the models constructed are generalized linear models or machine learning models.^{6,37–39} However, there are many kinds of scoring systems for disease progression in ALS,^{40–42} and the disease progression results of ALS patients may not be all the same, leading to different prognostic results. Moreover, dynamic prognostic outcomes also play an important role in clinical trials. Taking the time-to-event variable as the outcome is a more intuitive approach, and

the predicted survival probability can directly reflect the progress of the patient's condition. In addition, in studies with time-to-event as the outcome variable, traditional Cox PH models are often only used in univariate analysis or constructed multivariable prognostic models,^{12–16} which lack survival prediction ability at the individual level and do not take into account the dynamic effects of covariates. The ALS dynamic Cox model constructed in this paper cannot only identify dynamic effects in longitudinal covariates for prognostic analysis but also dynamically predict patients' future survival probability at the individual level.

When we apply the landmarking approach, some parameter settings are necessary. First, the prediction window w depends on the disease duration and the time interval of patient condition change.⁴ Therefore, predicting the survival of patients in the next 6 months can provide timely feedback on the condition of ALS patients. Second, the selection of the landmark time point s_l is independent of the actual survival time, which implies the weighting of the prediction time. The simplest method is to use an equidistant interval of points on the prediction interval $[0, s_{\max}]$, and the number of time points between 20 and 100 is appropriate.⁴³ In addition, the median follow-up time is usually selected as the maximum prediction time of interest s_{\max} . For example, in this article, 18 months is the intermediate of the maximum follow-up time (37.7 months) and is also the longest recorded stable period for ALS patients.²¹ The combination selection of prediction window w and landmark time points s_l can be obtained by cross-validation.⁴⁴ Finally, for the functional form of time-varying effects $\beta_k(s)$ and the baseline hazard changing $\theta(s)$, the most commonly used is quadratic functions.⁴⁵ Certainly, other functions could also be chosen to describe the change in time, such as the cubic function and spline function.

This study has some limitations. First, due to the high missing rate of raw data in PRO-ACT, some covariates were not included in the cleaned dataset, but they have been confirmed to affect ALS patients, such as the uric acid and creatine kinase concentrations.^{1,46} Second, the PRO-ACT database includes data from clinical trials of ALS patients over the past 20 years, although the ALS patients in the PRO-ACT database are not necessarily representative of the entire ALS patient population,⁴⁷ it is the largest publicly available dataset incorporating ALS clinical trials.¹ Although there are no genetics, cognitive status, or deep phenotype data in the PRO-ACT database, the clinical biomarkers of patients directly reflect the changes in patients' disease conditions. Moreover, the ALSFRS total score and its associated problem variables (Q1–Q10) were not statistically significant in the stepwise backward regression (except for Q8 and Q9, which were

statistically significant in the dynamic prediction model based on Dataset 3 but with small coefficients) and were not included in the final model. However, some studies have shown that, though functional decline based on the ALSFRS total score is closely related to survival, it does not perform better as a predictor.²⁶ Finally, although the model performed well in the internal cross-validation of Monte Carlo simulations, it should also be externally validated on datasets from other ALS centers to expand its application to more ALS patients. Although we demonstrated that the dynamic Cox model is a powerful tool, there is room for further improvement of the proposed dynamic prediction model as a reference for clinicians and patients.

Conclusions

In conclusion, this paper developed the first ALS dynamic survival prediction model based on the landmark approach in Cox regression by using ALS longitudinal clinical trial data from the PRO-ACT database. In addition to including baseline covariates, this model can also include longitudinal time-dependent covariates for analysis, which can identify the dynamic prognostic effect of these covariates and make individual survival predictions at any time point of interest for ALS patients with different conditions. The prognostic analysis and predictive results of this model will be valuable for improving the survival time of ALS patients and can provide a reference for clinicians to make clinical decisions about ALS disease treatment.

Author Contributions

Conception and design: Zheng Chen, Baoyi Huang; Acquisition of data: Zheng Chen; Analysis and interpretation of data: Baoyi Huang, Zheng Chen; Writing, review, and revision of the manuscript: Baoyi Huang, Zheng Chen, Xiang Geng, Zhiyin Yu, and Chengfeng Zhang; All authors provided final approval of the version to be published.

Acknowledgments

We thank the editor and reviewers for helpful comments on an earlier draft of the article. Data used in the preparation of this article were obtained from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT). As such, the following organizations and individuals within the PRO-ACT Consortium contributed to the design and implementation of the PRO-ACT Database and/or provided data, but did not participate in the analysis of the data or the writing of this report: ALS Therapy

Alliance, Cytokinetics, Inc., Amylyx Pharmaceuticals, Inc., Knopp Biosciences, Neuraltus Pharmaceuticals, Inc., Neurological Clinical Research Institute, MGH, Northeast ALS Consortium, Novartis, Prize4Life Israel, Regeneron Pharmaceuticals, Inc., Sanofi, Teva Pharmaceutical Industries, Ltd., The ALS Association.

Conflict of Interest

All authors have no relevant financial or non-financial interests to disclose.

Data Availability Statement

The datasets generated and/or analyzed during the current study are available in the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) repository. The dataset is provided by the PRO-ACT Consortium members and it is easily accessible after registration at the PRO-ACT website <https://ncr1.partners.org/ProACT>.

References

- Atassi N, Berry J, Shui A, et al. The PRO-ACT database: design, initial analyses, and predictive features. *Neurology*. 2014;83(19):1719-1725. doi:[10.1212/WNL.0000000000000951](https://doi.org/10.1212/WNL.0000000000000951)
- Kiernan MC, Vucic S, Talbot K, et al. Improving clinical trial outcomes in amyotrophic lateral sclerosis. *Nat Rev Neurol*. 2021;17(2):104-118. doi:[10.1038/s41582-020-00434-z](https://doi.org/10.1038/s41582-020-00434-z)
- Gola D, König IR. Empowering individual trait prediction using interactions for precision medicine. *BMC Bioinformatics*. 2021;22(1):74. doi:[10.1186/s12859-021-04011-z](https://doi.org/10.1186/s12859-021-04011-z)
- Westeneng HJ, Debray TPA, Visser AE, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol*. 2018;17(5):423-433. doi:[10.1016/S1474-4422\(18\)30089-9](https://doi.org/10.1016/S1474-4422(18)30089-9)
- Spencer KR, Foster ZW, Rauf NA, et al. Neuropathological profile of long-duration amyotrophic lateral sclerosis in military veterans. *Brain Pathol*. 2020;30(6):1028-1040. doi:[10.1111/bpa.12876](https://doi.org/10.1111/bpa.12876)
- Zandonà A, Vasta R, Chiò A, Di Camillo B. A dynamic Bayesian network model for the simulation of amyotrophic lateral sclerosis progression. *BMC Bioinformatics*. 2019;20(suppl. 4):118. doi:[10.1186/s12859-019-2692-x](https://doi.org/10.1186/s12859-019-2692-x)
- Lacomblez L, Bensimon G, Leigh PN, et al. Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis/Riluzole Study Group II. *Lancet*. 1996;347(9013):1425-1431. doi:[10.1016/s0140-6736\(96\)91680-3](https://doi.org/10.1016/s0140-6736(96)91680-3)

8. Abe K, Aoki M, Tsuji S, et al. Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* 2017;16(7):505-512. doi:10.1016/S1474-4422(17)30115-1
9. Zhou N, Manser P. Does including machine learning predictions in ALS clinical trial analysis improve statistical power? *Ann Clin Transl Neurol.* 2020;7(10):1756-1765. doi:10.1002/acn3.51140
10. Leão T, Madeira SC, Gromicho M, de Carvalho M, Carvalho AM. Learning dynamic Bayesian networks from time-dependent and time-independent data: unraveling disease progression in amyotrophic lateral sclerosis. *J Biomed Inform.* 2021;117:103730. doi:10.1016/j.jbi.2021.103730
11. Watanabe H, Atsuta N, Nakamura R, et al. Factors affecting longitudinal functional decline and survival in amyotrophic lateral sclerosis patients. *Amyotroph Lateral Scler Frontotemporal Degener.* 2015;16(3-4):230-236. doi:10.3109/21678421.2014.990036
12. Bald EM, Nance CS, Schultz JL. Melatonin may slow disease progression in amyotrophic lateral sclerosis: findings from the pooled resource open-access ALS clinic trials database. *Muscle Nerve.* 2021;63(4):572-576. doi:10.1002/mus.27168
13. Chen L, Zhang B, Chen R, et al. Natural history and clinical features of sporadic amyotrophic lateral sclerosis in China. *J Neurol Neurosurg Psychiatry.* 2015;86(10):1075-1081. doi:10.1136/jnnp-2015-310471
14. Wei Q, Chen X, Zheng Z, et al. The predictors of survival in Chinese amyotrophic lateral sclerosis patients. *Amyotroph Lateral Scler Frontotemporal Degener.* 2015;16(3-4):237-244. doi:10.3109/21678421.2014.993650
15. Marin B, Couratier P, Arcuti S, et al. Stratification of ALS patients' survival: a population-based study. *J Neurol.* 2016;263(1):100-111. doi:10.1007/s00415-015-7940-z
16. Reniers W, Schrooten M, Claeys KG, et al. Prognostic value of clinical and electrodiagnostic parameters at time of diagnosis in patients with amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener.* 2017;18(5-6):341-350. doi:10.1080/21678421.2017.1288254
17. Van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scand J Stat.* 2017;34(1):70-85. doi:10.1111/j.1467-9469.2006.00529.x
18. Yang Z, Hou Y, Lyu J, Liu D, Chen Z. Dynamic prediction and prognostic analysis of patients with cervical cancer: a landmarking analysis approach. *Ann Epidemiol.* 2020;44:45-51. doi:10.1016/j.annepidem.2020.01.009
19. Yang Z, Wu H, Hou Y, Yuan H, Chen Z. Dynamic prediction and analysis based on restricted mean survival time in survival analysis with nonproportional hazards. *Comput Methods Programs Biomed.* 2021;207:106155. doi:10.1016/j.cmpb.2021.106155
20. PRO-ACT Data Set. Accessed August 2022. <https://nctu.partners.org/PROACT>
21. Bedlack RS, Vaughan T, Wicks P, et al. How common are ALS plateaus and reversals? *Neurology.* 2016;86(9):808-812. doi:10.1212/WNL.0000000000002251
22. Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol.* 1972;34(2):187-202. doi:10.1111/j.2517-6161.1972.tb00899.x
23. Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr.* 2008;17(2):145-151. doi:10.1111/j.1466-8238.2007.00358.x
24. Gerds TA, Schumacher M. Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biom J.* 2006;48(6):1029-1040. doi:10.1002/bimj.200610301
25. Creemers H, Grupstra H, Nollet F, van den Berg LH, Beelen A. Prognostic factors for the course of functional status of patients with ALS: a systematic review. *J Neurol.* 2015;262(6):1407-1423. doi:10.1007/s00415-014-7564-8
26. Ong ML, Tan PF, Holbrook JD. Predicting functional decline and survival in amyotrophic lateral sclerosis. *PLoS ONE.* 2017;12(4):e0174925. doi:10.1371/journal.pone.0174925
27. Wei QQ, Chen Y, Cao B, et al. Blood hemoglobin A1c levels and amyotrophic lateral sclerosis survival. *Mol Neurodegener.* 2017;12(1):69. doi:10.1186/s13024-017-0211-y
28. McCombe PA, Henderson RD. Effects of gender in amyotrophic lateral sclerosis. *Gend Med.* 2010;7(6):557-570. doi:10.1016/j.genm.2010.11.010
29. Schumacher M, Hieke S, Ihorst G, Engelhardt M. Dynamic prediction: a challenge for biostatisticians, but greatly needed by patients, physicians and the public. *Biom J.* 2020;62(3):822-835. doi:10.1002/bimj.201800248
30. Tavazzi E, Daberdaku S, Zandonà A, et al. Predicting functional impairment trajectories in amyotrophic lateral sclerosis: a probabilistic, multifactorial model of disease progression. *J Neurol.* 2022;269(7):3858-3878. doi:10.1007/s00415-022-11022-0
31. Finnerup NB, Haroutounian S, Baron R, et al. Neuropathic pain clinical trials: factors associated with decreases in estimated drug efficacy. *Pain.* 2018;159(11):2339-2346. doi:10.1097/j.pain.0000000000001340
32. Ikeda K, Hirayama T, Takazawa T, Kawabe K, Iwasaki Y. Relationships between disease progression and serum levels of lipid, urate, creatinine and ferritin in Japanese patients with amyotrophic lateral sclerosis: a cross-sectional study. *Intern Med.* 2012;51(12):1501-1508. doi:10.2169/internalmedicine.51.7465
33. Povedano M, Paipa A, Barceló M, et al. Plasma exchange with albumin replacement and disease progression in amyotrophic lateral sclerosis: a pilot study. *Neurol Sci.* 2022;43(5):3211-3221. doi:10.1007/s10072-021-05723-z
34. Chiò A, Calvo A, Bovio G, et al. Amyotrophic lateral sclerosis outcome measures and the role of albumin and

- creatinine: a population-based study. *JAMA Neurol.* 2014;71(9):1134-1142. doi:[10.1001/jamaneurol.2014.1129](https://doi.org/10.1001/jamaneurol.2014.1129)
35. Grosskreutz J, Van Den Bosch L, Keller BU. Calcium dysregulation in amyotrophic lateral sclerosis. *Cell Calcium.* 2010;47(2):165-174. doi:[10.1016/j.ceca.2009.12.002](https://doi.org/10.1016/j.ceca.2009.12.002)
 36. Mandrioli J, Rosi E, Fini N, et al. Changes in routine laboratory tests and survival in amyotrophic lateral sclerosis. *Neurol Sci.* 2017;38(12):2177-2182. doi:[10.1007/s10072-017-3138-8](https://doi.org/10.1007/s10072-017-3138-8)
 37. Carreiro AV, Amaral PMT, Pinto S, Tomás P, de Carvalho M, Madeira SC. Prognostic models based on patient snapshots and time windows: predicting disease progression to assisted ventilation in amyotrophic lateral sclerosis. *J Biomed Inform.* 2015;58:133-144. doi:[10.1016/j.jbi.2015.09.021](https://doi.org/10.1016/j.jbi.2015.09.021)
 38. Pfohl SR, Kim RB, Coan GS, Mitchell CS. Unraveling the complexity of amyotrophic lateral sclerosis survival prediction. *Front Neuroinform.* 2018;12:36. doi:[10.3389/fninf.2018.00036](https://doi.org/10.3389/fninf.2018.00036)
 39. Pancotti C, Birolo G, Rollo C, et al. Deep learning methods to predict amyotrophic lateral sclerosis disease progression. *Sci Rep.* 2022;12(1):13738. doi:[10.1038/s41598-022-17805-9](https://doi.org/10.1038/s41598-022-17805-9)
 40. Balendra R, Jones A, Jivraj N, et al. Estimating clinical stage of amyotrophic lateral sclerosis from the ALS functional rating scale. *Amyotroph Lateral Scler Frontotemporal Degener.* 2014;15(3-4):279-284. doi:[10.3109/21678421.2014.897357](https://doi.org/10.3109/21678421.2014.897357)
 41. Chiò A, Hammond ER, Mora G, Bonito V, Filippini G. Development and evaluation of a clinical staging system for amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry.* 2015;86(1):38-44. doi:[10.1136/jnnp-2013-306589](https://doi.org/10.1136/jnnp-2013-306589)
 42. Lunetta C, Lizio A, Melazzini MG, Maestri E, Sansone VA. Amyotrophic lateral sclerosis survival score (ALS-SS): a simple scoring system for early prediction of patient survival. *Amyotroph Lateral Scler Frontotemporal Degener.* 2015;17(1-2):93-100. doi:[10.3109/21678421.2015.1083585](https://doi.org/10.3109/21678421.2015.1083585)
 43. Van Houwelingen HC, Putter H. *Dynamic Prediction in Clinical Survival Analysis.* CRC Press; 2012.
 44. Wu C, Li L, Li R. Dynamic prediction of competing risk events using landmark sub-distribution hazard model with multiple longitudinal biomarkers. *Stat Methods Med Res.* 2020;29(11):3179-3191. doi:[10.1177/0962280220921553](https://doi.org/10.1177/0962280220921553)
 45. Klein JP, van Houwelingen HC, Ibrahim JG, Scheike TH. *Handbook of Survival Analysis, Chapter 21. Landmarking.* Chapman & Hall/CRC Press; 2014:441-456.
 46. Rafiq MK, Lee E, Bradburn M, McDermott CJ, Shaw PJ. Creatine kinase enzyme level correlates positively with serum creatinine and lean body mass, and is a prognostic factor for survival in amyotrophic lateral sclerosis. *Eur J Neurol.* 2016;23(6):1071-1078. doi:[10.1111/ene.12995](https://doi.org/10.1111/ene.12995)
 47. Zach N, Ennist DL, Taylor AA, et al. Being PRO-ACTive: what can a clinical trial database reveal about ALS? *Neurotherapeutics.* 2015;12(2):417-423. doi:[10.1007/s13311-015-0336-z](https://doi.org/10.1007/s13311-015-0336-z)

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. The dynamic Cox model based on Dataset 3.

Table S2. The detailed covariate information for example patients.

Figure S1. Time-varying effects of covariates for the dynamic Cox model based on Dataset 3.

Figure S2. Flow chart of data cleaning.

Appendix S1. The process of data cleaning.